Evaluation of bone formation in calcium phosphate scaffolds with μCT – method validation using SEM

S. Lewin[1☯], A. Barba[2, 3, 4☯], C. Persson[1], J. Franch[2], MP. Ginebra[3, 4] and C. Öhman-Mägi[1*]

[1]Materials in Medicine Group, Div. of Applied Materials Science, Dept. of Engineering Sciences, Uppsala University, Lägerhyddsvägen 1, 75237 Uppsala, Sweden, [2]Bone Healing Group, Animal Medicine and Surgery Department, Veterinary School, Autonomous University of Barcelona (UAB), V Building, 08193 Bellaterra, Spain, [3]Biomaterials, Biomechanics and Tissue Engineering Group, Department of Materials Science and Metallurgical Engineering, Technical University of Catalonia (UPC), Av. Diagonal 647, Barcelona 08028, Spain, [4]Centre for Research in Nanoengineering, Technical University of Catalonia, C/ Pascuali Vila 15, Barcelona 08028, Spain.

☯These authors contributed equally to this work

*Corresponding author: Caroline.Ohman@angstrom.uu.se

Abstract
There is a plethora of Calcium phosphate (CaP) scaffolds used as synthetic substitutes to bone grafts. The scaffold performance is often evaluated from the quantity of bone formed within or in direct contact with the scaffold. Micro computed tomography (μCT) allows three-dimensional evaluation of bone formation inside scaffolds. However, the almost identical x-ray attenuation of CaP and bone obtrude the separation of these phases in μCT images. Commonly, segmentation of bone in μCT images is based on gray scale intensity, with manually determined global thresholds. However, image analysis methods, and methods for manual thresholding in particular, lack standardization and may consequently suffer from subjectivity. The aim of the present study was to provide a methodological framework for addressing these issues.

Bone formation in two types of CaP scaffold architectures (foamed and robocast), obtained from a larger animal study (a 12 week canine animal model) was evaluated by μCT. In addition, cross-sectional scanning electron microscopy (SEM) images were acquired as references to determine thresholds and to validate the result. μCT datasets were registered to the corresponding SEM reference. Global thresholds were then determined by quantitatively correlating the different area fractions in the μCT image, towards the area fractions in the corresponding SEM image. For comparison, area fractions were also quantified using global thresholds determined manually by two different approaches. In the validation the manually determined thresholds resulted in large average errors in area fraction (up to 17%), whereas for the evaluation using SEM references, the errors were estimated to be less than 3%. Furthermore, it was found that basing the thresholds on one single SEM reference gave lower errors than determining them manually.

This study provides an objective, robust and less error prone method to determine global thresholds for the evaluation of bone formation in CaP scaffolds.

Keywords: Calcium phosphate scaffolds, micro computed tomography, bone formation, image analysis and segmentation.

1. Introduction

Bone formation in CaP scaffolds is influenced by various factors including chemical composition, porosity and architecture of the scaffolds (Habraken *et al* 2016). The in vivo performances of scaffolds are generally evaluated by the quantity of bone formed within or in direct contact with the scaffold. Histology and immunohistochemical techniques have been the gold standards for evaluation of bone formation around implants or implant-bone interfaces (Appel *et al* 2013). These traditional techniques give a high resolution but have drawbacks such as being destructive, require a high workload and are two-dimensional (Appel *et al* 2013). Micro computed tomography (μCT) partly overcomes these shortcomings. Nonetheless, the image analysis methods used to quantify bone formation in CaP scaffolds from μCT images are not standardized and can suffer from subjectivity (Guldberg *et al* 2008, Peyrin 2011).

Quantitative evaluation of μCT images requires the images to be segmented. Segmentation approaches have been extensively discussed for characterization of bone and porous material (Bouxsein *et al* 2010, Ding *et al* 1999, Hara *et al* 2002, Kerckhofs *et al* 2008, Parkinson *et al* 2008, Particelli *et al* 2012, Perilli *et al* 2007). The selection of global thresholds in the grayscale intensity histogram is the most effective segmentation method and is considered robust for materials with well separated attenuation, such as evaluation of bone or CaP scaffolds alone (Bouxsein *et al* 2010, Campbell and Sophocleous 2014). However, the method has limitations when evaluating bone formation in CaP scaffolds due to the two materials similar composition and hence similar x-ray attenuation (Guldberg *et al* 2008, Jones *et al* 2007, Peyrin 2011). The limited resolution and noise from polychromatic beams in desktop-μCT further complicates the segmentation.

Relatively few studies have evaluated bone formation in CaP scaffolds by desktop-μCT (Gauthier *et al* 2005, Ishack *et al* 2017, Jones *et al* 2007, Lovati *et al* 2016, Polak *et al* 2012, Roberts *et al* 2011, Subramaniam *et al* 2016, Thimm *et al* 2013). One approach has been to segment the images by global thresholding and visually compare grayscale and thresholded μCT cross-sections (Ishack *et al* 2017, Roberts *et al* 2011). Guldberg *et al* (2008) argued that for new μCT applications it is advisable to compare thresholded 2D μCT-slices to registered histological sections. In some studies, histology is used together with μCT and the results are presented complementary (Ishack *et al* 2017, Lovati *et al* 2016, Subramaniam *et al* 2016), or to validate the μCT result (Gauthier *et al* 2005, Roberts *et al* 2011, Stalder *et al* 2014). Nevertheless, in most studies setting the thresholds included visual assessment. The methodology consequently is at risk of suffering from subjectivity since the different phases can be difficult to distinguish in desktop-μCT images.

More sophisticated algorithms for segmentation have been evaluated. Jones *et al* (2007) studied bone ingrowth in CaP scaffolds by a three phase segmentation approach that uses a multistep filter with anisotropic diffusion and edge enhancement, followed by an active contours algorithm (Sheppard *et al* 2004). The segmentation method was reported to work accurately, however the author specified that setting the initial thresholds included a level of subjectivity. Polak *et al* (2012) developed an automated segmentation algorithm to segment bone and CaP in μCT images. The automation removes the subjectivity, adds reproducibility, and is more time-efficient than manual evaluation. Nevertheless, the method assumes a regular lattice structure and cannot be used in irregular scaffolds, such as foamed scaffolds. Additionally, this method requires an initial threshold to differentiate between bone and scaffold. This threshold was optimized by correlating the result of the automated segmentation to results

of a manual segmentation. However, the similar attenuation of bone and CaP can impede the manual segmentation. Consequently, a separate validation would be needed to verify the results.

In order to objectively and in a reproducible way separate CaP scaffold from bone a new methodology is required. Either a fully automatic algorithm, non-sensitive to the initial threshold between bone and scaffold, or a standardized method to set thresholds. Using an external method as reference for quantitatively determining the thresholds has been used in porous materials (Kerckhofs *et al* 2008), trabecular and cortical bone (Ding *et al* 1999, Particelli *et al* 2012, Perilli *et al* 2007). One study was found using this approach to segment CaP scaffold and bone: an in vitro study Thimm *et al* (2013) used histology to set the thresholds for segmentation of μCT images of β-TCP scaffolds. The global thresholds were adjusted to obtain optimal correlations to parameters related to porosity and mineralization in the histological section. Mean percentage differences between histomorphometry and μCT measurements ranged from 1.4% (pore vertical diameter) to 14.0% (area of mineralized tissue). It was reported that the staining histology and light microscopy did not give an optimal resolution, and the possibility to use scanning electron microscopy (SEM) was suggested.

The main purpose of the present study was to develop and validate a method for evaluating desktop-μCT images of bone and CaP scaffolds in a more reproducible and objective manner than by manual thresholding. Two types of scaffold architectures were evaluated, i.e. fabricated by foaming or robocasting. The scaffolds were implanted intraosseously or intramuscularly in a canine animal model for 12 weeks. SEM images were used as references to determine thresholds and validate the result.

2. Materials & Methods

2.1. Preparation of the scaffolds

Cylindrical calcium deficient hydroxyapatite scaffolds (5mm diameter and 10mm height) were manufactured either by robocasting or by foaming, following protocols described elsewhere (Maazouz *et al* 2014, Montufar *et al* 2010). Examples of the two architectures can be seen in Figure 1.

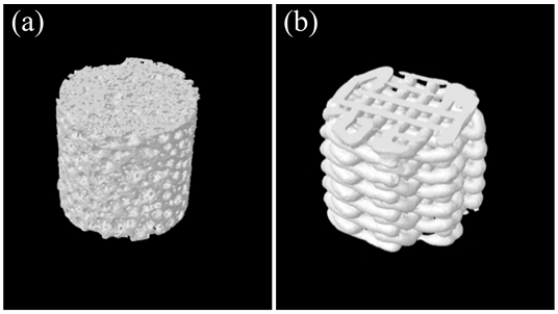

Figure 1. μCT 3D images of one foamed (a) and one robocast (b) scaffold. The diameter of the scaffolds is approximately 5 mm. The full height of the samples is 10 mm, here approximately half the height is shown.

2.2 In vivo model

The study was performed on three adult beagle dogs (body weight 14-17 kg) purchased from a professional stock breeder (Isoquimen S.L., Spain). All animal procedures in this study were in compliance with the Guide for Care and Use of Laboratory Animals (1996) and the European Community Guidelines (Directive 2010/63/EU) for the protection of animals used for scientific purposes (2010). Ethical approval for the procedures was obtained from the local Ethic Committee (CEAAH 2338). Besides

the intramuscular and intraosseous implantations described in the present study, the animals were used for a larger in vivo study where the performance of the scaffolds was further evaluated (results to be published separately).

Before the surgical procedure, dogs were preanesthetized using an intramuscular injection of medetomidine and methadone. Anesthesia was induced by an intravenous injection of propofol and diazepam, and was maintained with inhaled isoflurane in an oxygen carrier. For the intramuscular (IM) implantation, animals were placed in sternal recumbency and the lumbar areas were shaved and scrubbed with a 4% chlorhexidine gluconate solution for an aseptic preparation of the surgical field. Subsequently, one skin incision was performed on the lumbar region and a bilateral fascia incision was created in the paraspinal muscle. Using blunt dissection, intramuscular pockets were created on each side of the spine. Each intramuscular pocket was filled with one of the above mentioned cylindrical scaffolds. Closure of the surgical wound was performed routinely in layers. For the intraosseous (IO) implantation, animals were placed in lateral recumbency and both hind limbs were shaved and scrubbed with a 4 % chlorhexidine gluconate solution. Thereafter, a longitudinal incision was made and the lateral side of the femur was approached by blunt dissection. Circular monocortical bone defects (Ø=5 mm) were drilled under continuous irrigation of physiological saline. One of the cylindrical scaffolds was inserted in each defect by press-fit and the surgical wound was closed in layers. A rotatory allocation system and a block design were used. One scaffold of each architecture (foamed or robocast) was implanted intramuscularly and intraosseously in each dog, resulting in two IM and two IO scaffolds implanted per animal. After surgery, the animals subcutaneously received a prophylactic long-acting (14 days) antibiotic and pain relief by methadone and meloxicam. Animals were allowed to bear their full weight and received normal diet immediately after surgery. The animals were euthanized at 12 weeks post-implantation by an intravenous injection of pentobarbital sodium, after a pre-euthanasia sedation of medetomidine.

2.3. Sample harvest and histological processing

Immediately after euthanasia, the IM and IO samples were harvested and fixed in a 4% neutral buffered formalin (pH 6.9%) solution for 72 hours. For the IO samples a transverse section of the femur (including the implant) was retrieved, fixed and µCT scanned, see Figure 2. After fixation (and scanning in the case of IO samples), all samples were dehydrated in an increasing series of ethanol solutions and embedded in four different graded mixtures of ethanol and methyl methacrylate resin under vacuum conditions. Following the embedding the specimens were photo polymerized, obtaining poly (methyl methacrylate) (PMMA) blocks that in the case of IM samples were analyzed by µCT. After µCT scanning, IM samples were sectioned transversally (perpendicular to the longitudinal axis of the implant) into two equal pieces. The IO samples were divided longitudinally (along the longitudinal axis of the implant) and transverse to the femur in order to evaluate the full thickness of the cortical bone defect and the role of the periosteum and endosteum on the bone healing. One half of each sample was polished (EXAKT Cutting & Grinding System, EXAKT Corp., Norderstedt, Germany) prior to sputtering of the surface with carbon to allow for proper electrical conductivity for backscattered scanning electron microscopy analysis (BS-SEM). The other half was used for staining histology evaluations (results to be published in a parallel study).

2.4. µCT scanning procedure

All implants were µCT scanned (Skyscan 1172, Bruker microCT, Kontich, Belgium) after retrieval. The following settings were used; tube voltage 90 kV and current 112 µA. An Al-Cu filter was used during the scanning. The isotropic pixel size was 5 µm (IM samples) or 10 µm (IO samples). For all samples

corresponding cross-sections were reconstructed using the NRecon software (Bruker microCT, Kontich, Belgium), and saved as 8-bit gray-level images.

2.5. SEM images

The BS-SEM imaging (Zeiss Neon40 EsB CrossBeam, Zeiss, Germany) was performed at 20 keV and a magnification of 250X, using an automatic stitching system (SmartStitch, Zeiss, Germany), which automatically stitches acquired images together to form one micrograph of the entire sample.

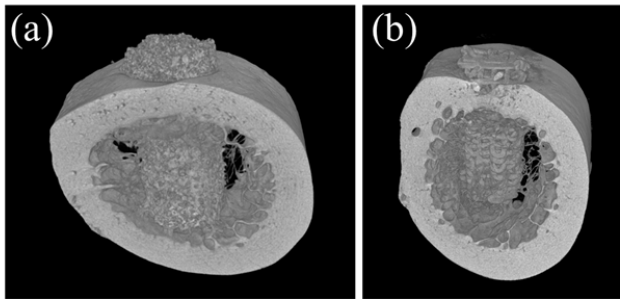

Figure 2. Reconstructed µCT scans of sectioned femurs with one foamed (a) and one robocast (b) implant inside the defect, after 12 weeks in vivo.

2.6. µCT image analysis: segmentation

All image processing and quantification was done using the CTAn software (Bruker microCT, Kontich, Belgium). The µCT datasets were segmented into three phases, to quantify the fractions of pore (air, soft tissue and, PMMA in IM samples), bone and scaffold. Two global thresholds were set for each sample. The lower threshold separates pore and bone, the upper threshold bone and scaffold. Three different methods for determining the global thresholds were evaluated, described in section 2.6.2.

2.6.1. Registration and region of interest (ROI)

Prior to segmentation, each µCT dataset was manually registered to its corresponding SEM image: the µCT dataset was rotated until one µCT cross-section matched the corresponding SEM section of that sample. The registration was made in the Dataviewer software (Bruker microCT, Kontich, Belgium). The same region of interest (ROI) could then be placed in the SEM image and in the corresponding µCT image.

The placement of the ROIs can be seen in Figure 3. In the images of the IO samples a polygonal ROI was placed in the plane along the longitudinal axis of the implant, transverse to the longitudinal axis of the femur, i.e. in the same direction as the SEM image was obtained. The ROI was positioned inside the defect, and within the cortical cortex. In the IM samples the SEM sections were obtained in the cross-sectional plane perpendicular to the longitudinal axis of the sample. A circular ROI was placed in this plane in the µCT and the SEM images of the IM samples. The described ROIs were used in two of the methods when determining the thresholds (Method 2 and Method 3 in section 2.6.2). Furthermore, these ROIs were used in the validation (section 2.8). Additionally, in Method 3 volumes of interest (VOIs) were used to set the thresholds. A VOI was obtained for each sample by interpolating a stack of ROIs throughout the 3D µCT dataset. In the IO samples the stack of polygonal ROIs were adjusted to fit the circular defect in the femur, and in the IM samples the stack of circular ROIs resulted in a cylindrical VOI along the longitudinal axis of the samples.

In the 3D quantification, VOIs consisted of a stack of circular ROIs placed in the cross-sectional plane perpendicular to the longitudinal axis of the sample, as previously described for the IM. For the IO samples a stack of the circular ROIs were positioned inside the defect. The number of stacked ROIs corresponded to the height of the cortical cortex.

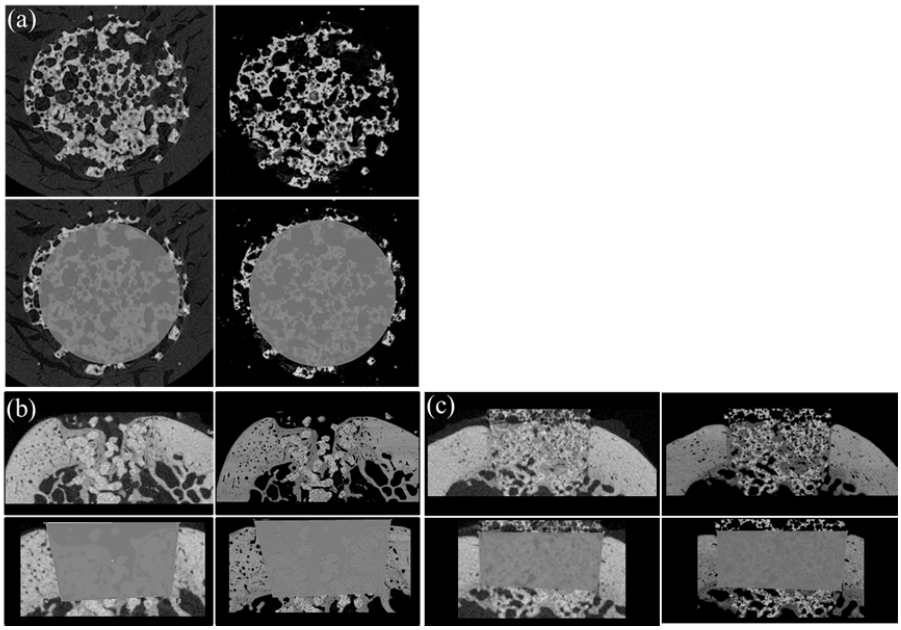

Figure 3. The placement of the ROIs for one IM foamed (a), one IO robocast (b) and one IO foamed (c) sample. In the upper row µCT (left) and corresponding SEM (right) image are shown for each sample. In the lower row for the IM sample a circular ROI is shown in gray. In the IO samples a polygonal ROI is shown in gray.

2.6.2. Thresholding methods

Method 1: manually determined thresholds
In the first method, global thresholds were determined manually to segment the three phases (pore, bone and scaffold). The thresholds were determined in the full three-dimensional dataset. The histograms and cross-sections for the whole dataset were visually inspected in order to separate the three phases.

Method 2: manually determined thresholds based on SEM reference images
In the second method, the thresholds were set for each sample in the 2D µCT cross-section from the registered dataset, using the corresponding SEM image as reference. The same ROI was selected in the SEM image and in the µCT image as previously described. The bone area was segmented in both the 2D µCT image and the corresponding SEM image. The operator compared the 2D µCT image and the SEM image to determine the global thresholds for separation of pore, bone and scaffold. The operator mainly focused on having the same amount of bone in the segmented µCT image as in the SEM image.

Method 3: automated thresholds based on SEM reference images
In the third method, the thresholds were set by quantitatively comparing the 2D µCT image and the corresponding SEM image.

In the SEM images the pore (PA_{SEM}), bone (BA_{SEM}) and scaffold (SA_{SEM}) area fractions were quantified by a 2D analysis. The high resolution of the SEM images made the three phases easily separable and

quantifiable. The same ROI was selected in the SEM image and in the corresponding 2D µCT cross-section, as in Method 2. However, a VOI was created in the full 3D µCT dataset by interpolation between the cross-sections. Segmentation by thresholding into pore, bone and scaffold fractions, and image processing was performed on the whole 3D µCT dataset. However, quantification of the pore ($PA_{\mu CT}$), bone ($BA_{\mu CT}$) and scaffold ($SA_{\mu CT}$) area fraction for each sample was only done on the registered 2D µCT cross-section, to be able to compare the area fractions from the µCT analysis to the area fractions quantified in the SEM image.

Figure 4 presents a flowchart illustrating the process to determine the thresholds in the µCT images. First the lower threshold, denoted as Th_1, was set by an iterative process where the fractions of pores were compared in the µCT image and in the SEM image. A low initial value for Th_1 in the µCT dataset was set to segment the pore fraction, using threshold: 0 to Th_1. After segmentation and image processing in 3D, a 2D analysis was made on the registered slice, to quantify the $PA_{\mu CT}$. The $PA_{\mu CT}$ was compared to the PA_{SEM}. Then Th_1 was increased, the dataset segmented, and the 2D analysis was repeated. Several iterations were made until the $PA_{\mu CT}$ was as close as possible to the PA_{SEM}. The same procedure was done for the bone fraction to set the upper threshold, denoted as Th_2. The bone fraction was segmented in the µCT dataset, using threshold: Th_1+1 to Th_2. Then a 2D analysis was made to quantify the $BA_{\mu CT}$ in the registered slice, and the $BA_{\mu CT}$ was compared to the BA_{SEM}. Then Th_2 was increased, the dataset segmented, and the 2D analysis was repeated. Several iterations were made until the $BA_{\mu CT}$ was as close as possible to the BA_{SEM}. The obtained thresholds, Th_1 and Th_2, for each sample were considered to be optimized. The µCT scaffold area fraction was quantified as the remaining area fraction in the µCT image: $SA_{\mu CT} = 1 - PA_{\mu CT} - BA_{\mu CT}$.

Since the grayscales for bone and scaffold are overlapping, some image processing was used to remove the scaffold marked as bone. The image processing was done using implemented functions in CTAn and consisted of morphological opening followed by despeckling. The same image processing for all samples of the same type was used when determining the thresholds and when doing the 3D quantification.

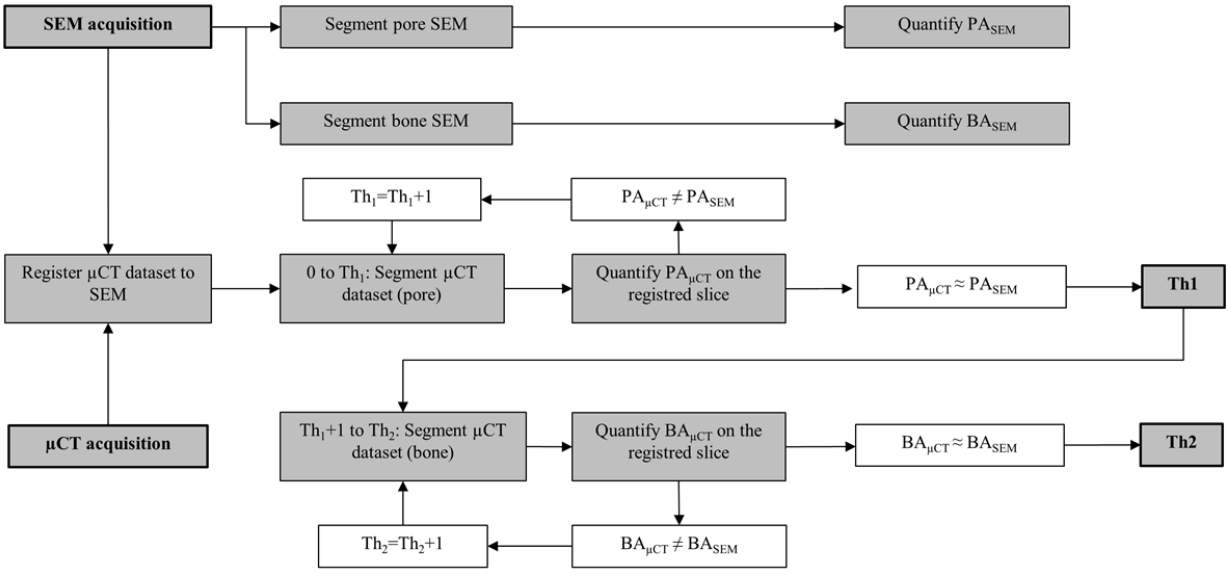

Figure 4. Flowchart illustrating the process to determine thresholds using Method 3.

2.7. Comparison of the different methods to the SEM results

The thresholds resulting from the three different methods were examined for each sample by comparing the pore, bone and scaffold area fraction in each SEM image, to the corresponding area fraction quantified in the registered 2D μCT image. The area fraction of all phases was quantified in the SEM images by a 2D analysis, as described in section 2.6.2 for Method 3. For the μCT cross-section, segmentation and image processing were made in the whole μCT 3D dataset, while the area fractions were quantified in the registered μCT image in a 2D analysis. The absolute difference in area fractions in the μCT image compared to the SEM was determined for each sample. The mean value of the absolute difference was calculated for each type of sample and for each phase as:

$$\begin{aligned}
\overline{\Delta PA} &= \frac{\sum_1^n \left| PA_{SEM} - PA_{\mu CT} \right|}{n}, \\
\overline{\Delta BA} &= \frac{\sum_1^n \left| BA_{SEM} - BA_{\mu CT} \right|}{n}, \\
\overline{\Delta SA} &= \frac{\sum_1^n \left| SA_{SEM} - SA_{\mu CT} \right|}{n},
\end{aligned} \tag{1}$$

where $PA_{SEM,}$ BA_{SEM} and SA_{SEM} are the pore, bone and scaffold area fraction obtained from the SEM image, $PA_{\mu CT,}$ $BA_{\mu CT}$ and $SA_{\mu CT}$ denote the pore, bone and scaffold area fraction obtained from the μCT image, and n denotes the number of samples.

2.8. Validation on the IM samples

Since the SEM images had been used when determining the thresholds, these could not be further used to validate Method 2 and 3. Therefore, on the IM foamed samples ($n=3$) a second transversal sectioning was made and analyzed by BS-SEM. The BS-SEM analysis was conducted according to section 2.5, resulting in previously unused SEM images, which could be registered to the μCT dataset and used for validation. In the additional SEM images, area fractions were quantified as described above. The thresholds determined for each method in section 2.6.2 were applied, respectively, to the 3D μCT dataset and a 2D analysis was made on the second registered cross-section for each sample. The errors were computed for each method, each type of sample, and each phase, as the mean value of the absolute difference, as described in section 2.7.

2.9. Using SEM-reference samples for evaluation of a larger dataset

The next part of the study was done to assess the possibility to use less SEM reference samples to set thresholds for a larger number of samples. This part was also used to test the sensitivity of the thresholds.

First an evaluation was made using two SEM reference samples. The mean value of the optimized thresholds for two of the samples, both lower and upper, was used on the third sample. The pore, bone and scaffold area fractions were quantified in 2D on the registered cross-section for each sample, as described in section 2.7. This was repeated for all samples of each type (n=3). The absolute difference in area fractions in the μCT image compared to the SEM was determined for each sample and the mean values for each sample type and each phase were calculated using equation (1) in section 2.7.

The same procedure was done using only one SEM reference sample. The optimized threshold from one sample was applied in the analysis of the other two samples, giving six evaluations for each sample type (n=6). 2D area fractions and absolute difference in area fractions were determined as described in the paragraph above.

2.10. Quantification and visualization in 3D

The previously established thresholds from the different methods were used to quantify bone, air and scaffold volume fractions. Surface rendered 3D models were created of the bone and scaffold in the samples. The 3D models were visualized using the CTVol software (Bruker microCT, Kontich, Belgium).

3. Results

The IM robocast samples did not show any bone formation, hence, the analysis of these samples was limited to the fraction of scaffold and pore. Therefore, no SEM references were obtained and the results were excluded.

3.1. µCT image analysis: segmentation

Examples of gray-level intensity histogram from the different types of samples can be seen in Figure 5. The figure shows the thresholds for each sample using the different methods for thresholding. The lower and upper thresholds are denoted as Th_1 and Th_2, respectively. Due to the similar attenuation of bone and the CaP scaffold, these phases overlap in the histograms. In the IM foamed and the IO robocast samples bimodal distributions were seen between the pore and bone-scaffold phases. The IO foamed sample did not show a clear bimodal distribution between the pore and bone-scaffold phases. For each type of sample one SEM and one µCT image can be seen in Figure 6, together with the corresponding segmented image (using thresholds from Method 3).

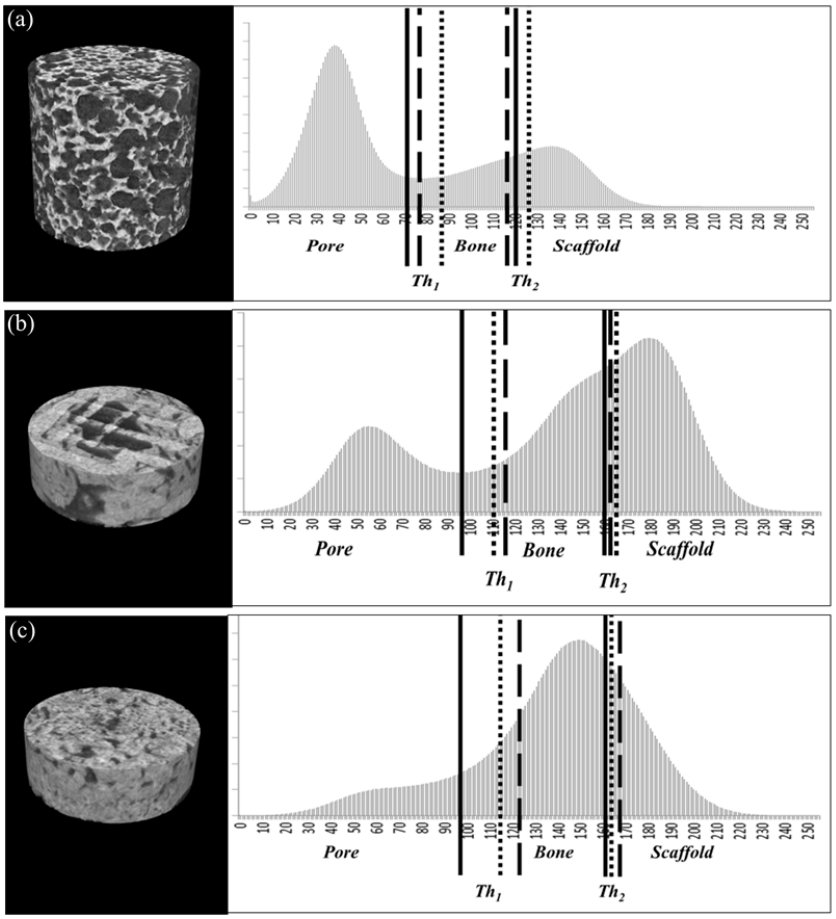

Figure 5. 3D µCT reconstruction and grayscale intensity histogram of the VOIs of one IM foamed sample (a), one IO robocast sample (b) and one IO foamed sample (c). In the histogram the thresholds are marked as Th_1 and Th_2. The solid line represents the thresholds from Method 1, the dashed line from Method 2 and the dotted line from Method 3.

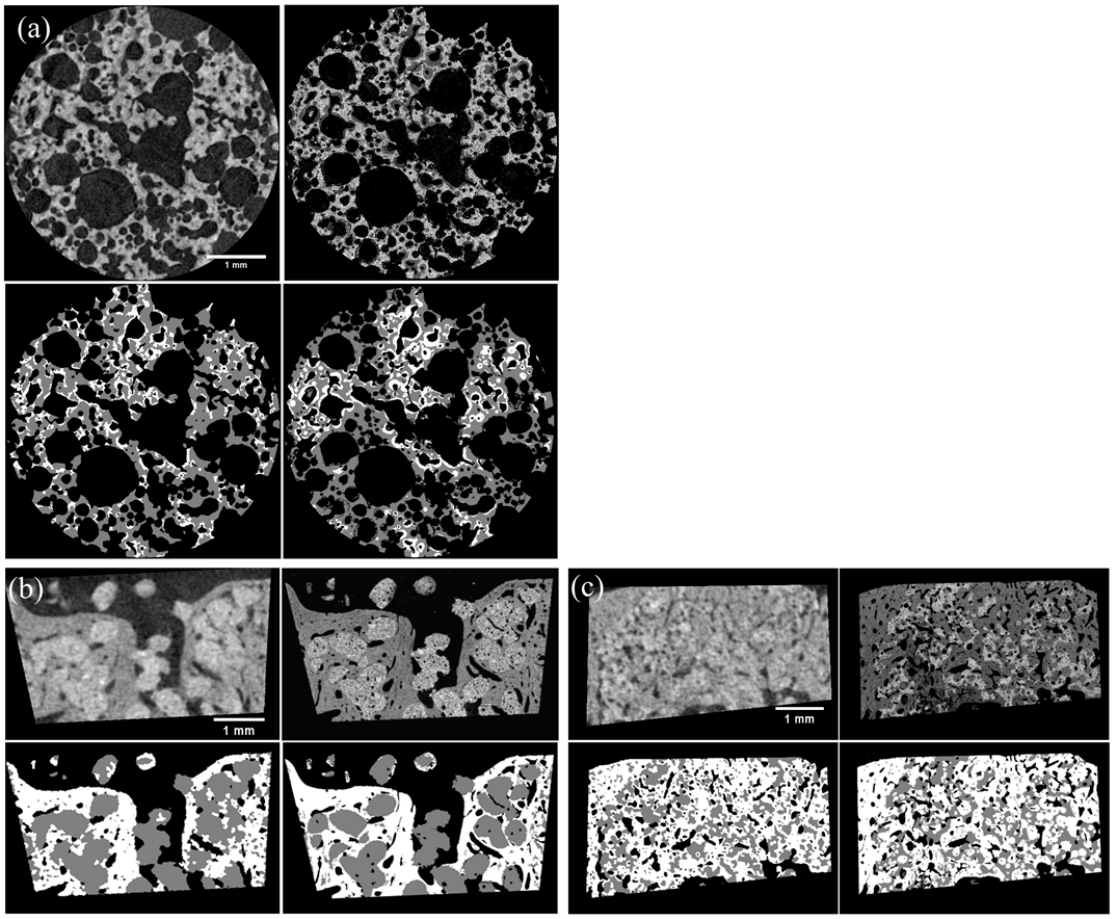

Figure 6. Representative µCT cross-sections (left) and corresponding SEM images (right) after 12 weeks; one IM foamed sample (a), one IO robocast sample (b) and one IO foamed sample (c). Below the µCT and SEM images are the respective segmented images with bone in white and remaining scaffold in gray. The images were segmented using Method 3.

3.2. Comparison of the different methods to the SEM results

Mean values of pore, bone and scaffold area fractions from the µCT and the SEM images can be seen in Table 1, for each type of sample and each method. The mean values of the absolute difference for each phase ($\overline{\Delta PA}$, $\overline{\Delta BA}$ and $\overline{\Delta SA}$) are presented in Figure 7. Method 1 (manually determined thresholds) gave mean area fractions with relatively large differences to the result from the SEM images for all types of samples and all phases. The pore area fraction was underestimated for all types of samples (maximum $\overline{\Delta PA}$ was 7% for the foamed IO samples). The bone area fraction was underestimated for the IO robocast samples, while the IO foamed and IM foamed samples was overestimated (maximum $\overline{\Delta BA}$ was 13% for the IM foamed samples). The scaffold area fraction was overestimated for IO robocast and IO foamed samples, but underestimated for the IM foamed samples (maximum $\overline{\Delta SA}$ was 7% for the IO robocast samples).

Using Method 2 (manually determined thresholds based on SEM reference images) the difference between the result in the SEM and the µCT decreased. The maximum $\overline{\Delta PA}$ of 4% was found in the IO foamed and

IM foamed samples. The maximum $\overline{\Delta BA}$ and $\overline{\Delta SA}$ (4% and 8%, respectively) were found in the IO foamed samples.

For Method 3 (automated thresholds based on SEM reference images), the mean values of the absolute difference in pore, bone and scaffold area were less than 1% for all individual samples since the threshold had been optimized using these SEM images.

Table 1. Mean values and standard deviations of pore, bone and scaffold area fractions for the different types of samples (*n*=3) using the three different methods. The results from the SEM images are also included.

	Pore area fraction (%)			Bone area fraction (%)			Scaffold area fraction (%)		
	Robo IO	Foam IO	Foam IM	Robo IO	Foam IO	Foam IM	Robo IO	Foam IO	Foam IM
SEM	31±8	17±2	57±9	31±7	52±5	12±12	39±6	31±3	31±3
Method 1	25±9	9±1	52±6	30±6	56±5	23±7	45±3	35±4	25±1
Method 2	30±9	21±2	55±6	30±7	56±6	15±11	40±8	23±4	31±5
Method 3	31±8	17±2	57±9	31±7	53±5	12±12	38±6	30±3	31±4

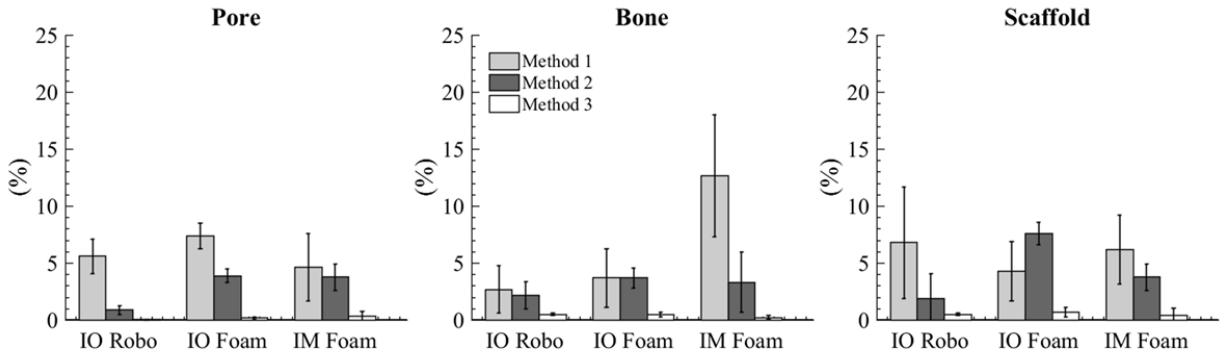

Figure 7. The absolute difference in pore, bone and scaffold area fractions comparing the SEM image to the corresponding 2D µCT image using the thresholds from Method 1, Method 2 and Method 3. The results are presented for each type of samples, as the mean value (*n*=3) with error bars corresponding to the standard deviation.

3.3. Validation on the IM samples

The errors (mean value of the absolute difference) obtained from the validation on the second set of SEM images (foamed IM samples only) are presented in Table 2. Generally, Method 3 gave the smallest errors, while Method 1 gave the largest errors. Method 2 was found to give larger errors than Method 3 for pore and bone area fractions. For the scaffold area fraction the errors were equally large, however Method 3 had a smaller standard deviation. The overall largest error was found for Method 1, 17±11% in bone area fraction. For Method 2 the largest error was also seen in bone area fraction (6±3%), while the largest error in Method 3 was seen in scaffold area fraction (3±2%).

Table 2. Errors from the validation for the three IM foamed samples at the second cross-sections. Errors are presented as mean values (*n*=3) together with the standard deviations.

	Error pore area fraction (%)	Error bone area fraction (%)	Error scaffold area fraction (%)
Method 1	6±1	17±11	11±10
Method 2	3±1	6±3	3±4
Method 3	2±2	2±2	3±2

3.4. Using SEM reference samples for evaluation of a larger dataset

To reduce the amount of destructive SEM references and to test the sensitivity of the thresholding an additional evaluation was made, investigating if thresholds optimized for a smaller number of samples could be used on a larger group of samples. The number of SEM references was either one or two for each material type.

Figure 8 presents the mean value of the absolute differences in 2D for each phase ($\overline{\Delta PA}$, $\overline{\Delta BA}$ and $\overline{\Delta SA}$) and each sample type, compared to the SEM results for each individual sample. The absolute difference slightly increased (with up to 2 %) when one instead of two SEM references was used. Using one reference sample, the maximum absolute difference found was 3% (in $\overline{\Delta SA}$) for the IO robocast samples, and 4% (in $\overline{\Delta BA}$ and $\overline{\Delta SA}$) for the IO foamed samples. For the IM foamed samples the difference in pore area fraction was similar to the maximum differences in the IO samples (3% in $\overline{\Delta PA}$), while larger differences were found in bone and scaffold area fraction (10% in both $\overline{\Delta BA}$ and $\overline{\Delta SA}$) using one SEM reference sample.

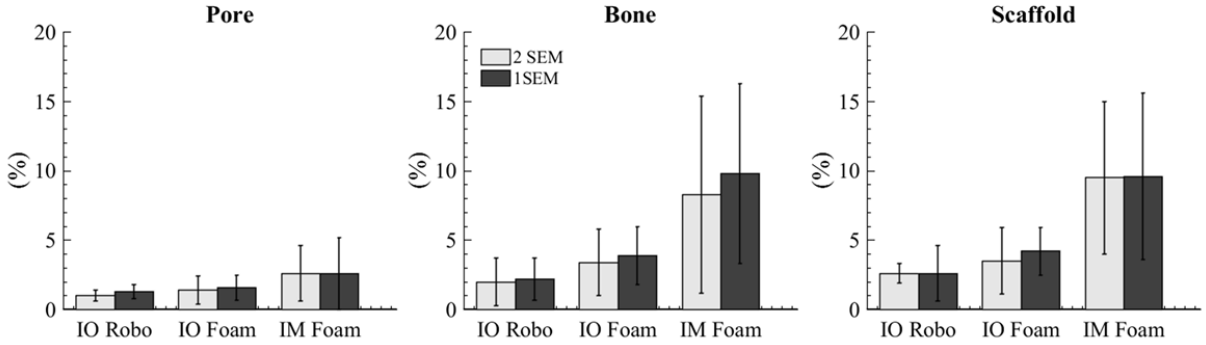

Figure 8. The absolute difference in pore, bone and scaffold area fractions comparing the SEM image to the corresponding 2D µCT image using thresholds based on one (*n*=6) or two (*n*=3) SEM reference samples. The results are presented for each type of samples, as the mean value with error bars corresponding to the standard deviation.

3.5. Quantification and visualization in 3D

The volume fractions were quantified in the 3D µCT datasets using the thresholds from the different methods (see Table 3). Comparing the results from Method 3 to the results using Method 1 the largest difference, 11%, was found in bone volume fraction for the IM foamed samples. Using Method 2 the largest difference compared to using Method 3 was found in scaffold volume fraction, 7%, for the IO foamed samples. The results from the analysis with one or two SEM references were similar to the results obtained using the thresholds from Method 3, and a maximum difference of 2% in volume fraction was found using both one and two SEM references.

Examples of 3D models for the three different types of samples, using the thresholds from Method 3, are visualized in Figure 9. The mean values of the upper and lower threshold for each type of sample are presented in table 4 to show the variation in threshold between the scaffold types using Method 3.

Table 3. Pore, bone and scaffold volume fractions for Method 1-3 (*n=3*) and the result from the evaluation using one (*n=6*) and two (*n=3*) SEM reference samples. The table displays the mean values and the standard deviation.

	Pore volume fraction (%)			Bone volume fraction (%)			Scaffold volume fraction (%)		
	Robo IO	Foam IO	Foam IM	Robo IO	Foam IO	Foam IM	Robo IO	Foam IO	Foam IM
Method 1	24±11	9±3	52±5	29±6	57±6	22±6	47±6	34±3	26±2
Method 2	29±10	21±5	55±5	26±6	56±9	13±9	45±8	23±5	32±4
Method 3	29±11	17±4	57±8	28 ±8	54±8	11±11	43±7	30±4	32±5
2 SEM	29±10	17±4	57±3	28±11	54±4	9±4	43±5	29±4	34±5
1 SEM	29±10	17±4	57±4	28±10	53±5	10±7	43±5	30±5	33±6

Table 4. Mean values and standard deviation of the upper and lower thresholds (using Method 3) for each sample type (*n=3*).

	Robo IO	Foam IO	Foam IM
Th_1	114±3	115±3	83±7
Th_2	165±2	163±2	118±8

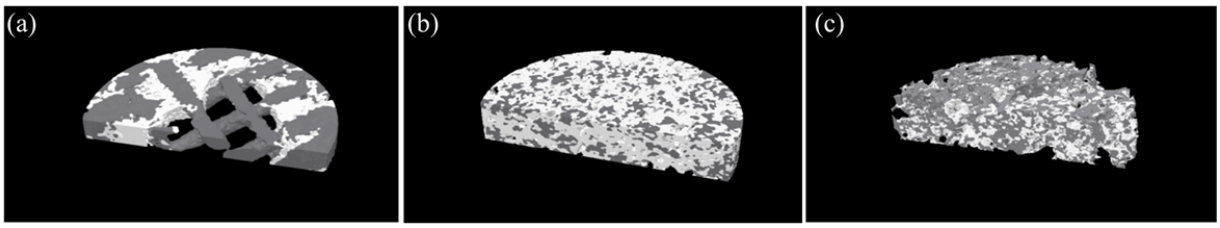

Figure 9. Surface rendered 3D models of an IO robocast sample (a), an IO foamed sample (b) and an IM foamed sample (c) after 12 weeks in vivo. A circular ROI (Ø=4 mm) is shown with the remaining scaffold in gray and the newly formed bone in white.

4. Discussion

The aim of this study was to develop an objective and reproducible method to determine thresholds for segmentation of bone and CaP scaffolds in µCT images, using SEM reference images. For comparison, the thresholds were also determined manually by two approaches. The segmentations were validated using another set of SEM images.

4.1. µCT image analysis: segmentation

The segmentation process in this study was complicated since bone and CaP scaffolds have similar x-ray attenuation. The complexity is further increased by scaffold degradation and bone mineralization. Consequently, the phases overlap in the intensity histogram, see Figure 5.

The obtained resolution, and the overlap in the histograms, did not allow automatic thresholding. Many automatic methods, such as the Otsu-method (Otsu 1979), often rely on bimodal distributions in the histograms. Adaptive thresholding methods also failed for the above-mentioned reasons. Therefore, global thresholding was used together with image processing. Due to the use of global thresholds, the segmentation does not account for errors, such as partial volume effects (PVE). PVE arises on object edges because the grayscale value of a voxel is altered depending on the proportion of materials with different attenuation inside the voxel (Ritman 2011). Consequently, some scaffold voxels were marked as bone and some bone marked as scaffold. A consistent image processing was used to remove some of the incorrectly marked voxels.

The images were manually registered to find the μCT cross-section corresponding to the SEM image. An accurate automated registration would have been preferred to save time. However, not many 2D-3D registration methods exist for such different resolutions. Furthermore, Stalder *et al* (2014) automatically registered histology images of bone and CaP biomaterials to synchrotron μCT images, but found that the manual registration gave more accurate results than the automated method.

4.2. Comparison of the different methods to the SEM results

In Method 1, the thresholds were set manually through visual evaluation of the images and the histograms. This method has commonly been used in studies for scaffold materials with an attenuation well separated from bone e.g. polymers (Guldberg *et al* 2008). Nevertheless, previous studies have pointed out the risk for subjectivity when thresholds are set to segment between bone and CaP scaffolds (Jones *et al* 2007, Peyrin 2011). This is a risk especially when, as in this study, desktop-μCT images with limited resolution are used. In the μCT images, the structures and phases were not visually distinguishable, especially not for the foamed scaffolds (see Figure 6). Consequently, using Method 1, the mean values of the absolute difference in area fractions, between the μCT cross-sections and the SEM images were relatively large (3-13%). In Method 2, the SEM image was used as a reference when setting the thresholds for the corresponding μCT cross-section. The mean values of the absolute difference in area fractions between μCT and SEM were lower (1-8%) in comparison to using Method 1. In Method 3 the thresholds were adjusted in an iterative process until the area fraction of each phase was as close as possible to the values from the SEM reference. Consequently, the mean values of the absolute differences between the μCT and the SEM results were small (<1%).

4.3. Validation on the IM samples

A second set of SEM images was obtained for validation from the IM foamed samples. The first set of SEM images could not be utilized since they had been used when determining the thresholds for Method 2 and Method 3. Again, the area fractions in a μCT cross-section were compared to the area fractions in the corresponding SEM image. The largest errors for Method 1 (17±11%) and Method 2 (6±3%) were seen in bone area fraction. For Method 3 the largest error of 3±2% was seen in the scaffold area fraction. As mentioned in the previous section the phases of bone and scaffolds were hard to distinguish visually in the μCT images, which explain the large errors using Method 1. The error decreased using Method 2, which demonstrates that reference images were needed. The error further decreased using Method 3, in which a quantitative and methodological approach was used.

In the in vitro study by Thimm *et al* (2013) staining histology was used to set the thresholds for segmentation of μCT images of human mesenchymal stem cells seeded β-TCP scaffolds. They adjusted the global threshold to obtain correlation to parameters related to porosity and mineralization in the

histological section. In the present study, the thresholds were instead adjusted to be optimized to the area fractions in SEM reference images, which provides an evaluation related to all three phases in the image. However, an exact correlation in area fractions in the SEM and μCT images might not signify that exactly the correct voxels are segmented as pore, bone and scaffold. The errors in the segmentation are in part related to partial volume effects due to the limited resolution of the μCT images. However, this type of errors will be present using global thresholding regardless of the method used to determine the thresholds, which further highlights the need of an objective method. Method 3 was considered more objective than the manual thresholding methods and also the results are reproducible. Furthermore, this method gave the lowest errors in the validation.

4.4. Using SEM reference samples for evaluation of a larger dataset

Using SEM references for thresholding removes the non-destructive element, one of the most important benefits of μCT. With this in mind another evaluation was made where the mean values of thresholds from a small number of samples were used on the rest of the samples of the same type. The aim was to reduce the amount of destructive SEM references needed for each type of sample, by finding a set of thresholds that could be used for a larger number of samples. This evaluation also indicates the sensitivity of setting the thresholds. For the IO samples the maximum mean absolute differences in area fraction was 4%, using only one reference for all three samples instead of one per sample. However, for the IM samples the maximum mean absolute difference in area fraction was up to 10%. Nevertheless, the mean differences were still smaller than when using Method 1, i.e. manual thresholding (up to 17%).

4.5. Quantification and visualization in 3D

The quantification in 3D showed maximum 10% and 8% difference using Method 1 and Method 2, compared to Method 3. Using one or two SEM references, the results in volume fraction differed maximum 2% from the results using Method 3. However, as we observed in the previous section, in 2D for individual samples, larger errors were observed especially for the IM foamed samples. This can be explained by looking at the variation in thresholds in Table 4 and the standard deviations of bone volume fractions in Table 3. In general, the IO samples of the respective material type (foamed or robocast) had a homogenous distribution of the phase fractions. Conversely, the amount of bone found in the IM samples varied considerably (1-22%) and as a consequence the variation in threshold of the IM samples was larger than for the IO samples (see Table 4). In the 3D quantification, this resulted in an overestimation of bone for some samples and an underestimation for others, therefore the difference in mean values in comparison to using Method 3 was still small.

To pre-calibrate the threshold on high-resolution SEM reference images on a smaller number of samples and use these in a larger number of samples allows for further mechanical or biological testing on the remaining samples. Nonetheless, using this approach it is important to maintain the same acquisition conditions (scanning parameters, sample embedding etc.). Thresholds should also be set for each type of samples i.e. same material, architecture, etc. However, it would not be recommended to use the mean value threshold approach if the bone formation is very different among the samples, as in the IM samples in this study. The limited number of samples in the present study does not allow for any recommendations on the exact number of reference samples which should be used in future studies. Nevertheless, it was found that basing the thresholds on one single reference sample gave lower errors than determining them manually. Furthermore, the use of one or two reference samples did not strongly influence the results of the IO samples, indicating that few samples could be enough, at least in samples with similar amounts of

bone. The use of an external method to pre-determine thresholds for segmentation has been implemented previously for both trabecular (Ding *et al* 1999, Perilli *et al* 2007) and cortical bone (Particelli *et al* 2012) and foamed materials (Kerckhofs *et al* 2008). The mentioned studies used external methods (i.e. Archimedes' principal, histology and optical microscopy) on a sub-set of samples to determine global thresholds which could thereafter be used on larger datasets.

Often μCT is used to replace histology and high-resolution 2D methods. A 3D evaluation gives additional and more reliable information than a 2D quantitative evaluation of only one or a few cross-sections. μCT allows 3D quantifications of bone formation, porosity, bone distribution and scaffold degradation. However, μCT will never completely replace evaluations by SEM or stained histology. These techniques provide different required information such as the presence of bone cells (osteoblasts, osteocytes and osteoclasts) and grade of bone tissue maturity, fibrous tissue infiltration, neovascularization and peri-implant inflammatory reaction. By using the method in the present study to set thresholds, μCT is not a replacement for histology, rather the two image modalities are used complementary. Hence, the SEM images give information on another scale as well as validate the segmentation of the μCT images. In future studies, scanning the scaffolds before implantation would be beneficial since scaffold degradation could be studied and the bone formation in each scaffold could be correlated to its architecture. Furthermore, following the bone ingrowth over time, through longitudinal in vivo μCT studies, would also be of interest.

Another approach to the problem in this paper would be to increase the radiopacity of the scaffolds. However, to the knowledge of the authors radopacifiers cannot be added without modifying the material properties of CaP, including both mechanical properties and biocompatibility. Further, radiopacifiers are generally non-degradable and their use in degradable CaP biomaterials is therefore limited (Lopez *et al* 2014).

The resolution of the μCT images obtained in this study made the segmentation of bone and scaffold complex. However, scanning with a higher resolution would have resulted in a smaller field of view, hence, only a part of the sample would have been scanned. For the IO samples the whole femur was scanned (10μm voxel size) to be able to evaluate the full thickness of the cortical shell. For the IM samples only the scaffold was scanned (5 μm voxel size). A higher resolution could facilitate the separation of bone and scaffold. With the use of a synchrotron source, it is possible that an accurate separation of bone and scaffold could have been done manually by visual examination of histograms and μCT cross-sections or by an automatic method. However, synchrotron sources are rare and not available for all studies (Britz *et al* 2010). Desktop-μCT is still the most widely used μCT technique and the present study contributes to meeting the need for improved methods for evaluation of desktop-μCT images.

5. Conclusions

A methodological framework for the evaluation of bone formation in CaP scaffolds was developed by using μCT combined with high-resolution SEM reference images to quantitatively determine the thresholds for segmentation. The results from using this method were compared to the results from two methods where the thresholds were set manually. Manually setting the thresholds in the μCT images gave large errors when the area fractions from the SEM histological images were compared to the area fractions from the μCT images. In the quantitative method employed in this study, the thresholds were determined by minimizing this error, which provided objectivity and robustness. The methods were validated with a second set of SEM images. The errors in area fractions were estimated for all phases to be less than 3% for

the method using quantitatively determined thresholds. The two methods where the thresholds were determined manually resulted in a less accurate evaluation, with errors up to 17% and 6%, respectively. In an additional evaluation, made to reduce the amount of destructive SEM references, the mean values of the thresholds for a smaller amount of samples were used for a larger number of samples. The approach resulted in differences in pore, bone and scaffold volume fractions comparable to using the thresholds determined for each samples. However, the difference was larger for the IM samples, which had a greater variation in the amount of bone formed in the samples.

In summary, to be able to characterize bone formation in CaP scaffolds in μCT images there is a need to standardize this type of evaluations, in order to allow for reproduction and comparison of different studies. While the present study represents a step forward, it is important that methods for segmentation of desktop-μCT images, required to quantify bone formation in CaP scaffolds, are further improved.

Acknowledgements
The authors are grateful for funding from the Swedish Foundation for International Cooperation in Research and Higher Education (STINT), project IG2011-2047, and from the Spanish Government, project MAT2015-65601-R co-funded by the EU through European Regional Development Funds. AB acknowledges a FPU scholarship from the Spanish Ministry of Education and MPG the ICREA Academia award from the Generalitat de Catalunya.

6. References

Appel A A, Anastasio M A, Larson J C and Brey E M 2013 Imaging challenges in biomaterials and tissue engineering *Biomaterials* **34** 6615-30

Bose S, Roy M and Bandyopadhyay A 2012 Recent advances in bone tissue engineering scaffolds *Trends Biotechnol.* **30** 546-54

Bouxsein M L, Boyd S K, Christiansen B A, Guldberg R E, Jepsen K J and Müller R 2010 Guidelines for assessment of bone microstructure in rodents using micro-computed tomography *J. Bone Miner. Res.* **25** 1468-86

Britz H m., Jokihaara J, Leppänen O v., Järvinen T and Cooper D m. l. 2010 3D visualization and quantification of rat cortical bone porosity using a desktop micro-CT system: a case study in the tibia *J. Microsc.* **240** 32-7

Campbell G M and Sophocleous A 2014 Quantitative analysis of bone and soft tissue by micro-computed tomography: applications to ex vivo and in vivo studies *BoneKEy Rep.* **3** 564

Ding M, Odgaard A and Hvid I 1999 Accuracy of cancellous bone volume fraction measured by micro-CT scanning *J. Biomech.* **32** 323-6

European Community Guidelines 2010 Directive 2010/63/EU of the European Parliament and of the Council of 22 September 2010 on the protection of animals used for scientific purposes

Gauthier O, Muller R, Vonstechow D, Lamy B, Weiss P, Bouler J, Aguado E and Daculsi G 2005 In vivo bone regeneration with injectable calcium phosphate biomaterial: A three-dimensional micro-computed tomographic, biomechanical and SEM study *Biomaterials* **26** 5444-53

Guldberg R E, Duvall C L, Peister A, Oest M E, Lin A S P, Palmer A W and Levenston M E 2008 3D Imaging of Tissue Integration with Porous Biomaterials *Biomaterials* **29** 3757-61

Habraken W, Habibovic P, Epple M and Bohner M 2016 Calcium phosphates in biomedical applications: materials for the future? *Mater. Today* **19** 69-87

Hara T, Tanck E, Homminga J and Huiskes R 2002 The influence of microcomputed tomography threshold variations on the assessment of structural and mechanical trabecular bone properties *Bone* **31** 107-9

Ho S T and Hutmacher D W 2006 A comparison of micro CT with other techniques used in the characterization of scaffolds *Biomaterials* **27** 1362-76

Ishack S, Mediero A, Wilder T, Ricci J L and Cronstein B N 2017 Bone regeneration in critical bone defects using three-dimensionally printed β-tricalcium phosphate/hydroxyapatite scaffolds is enhanced by coating scaffolds with either dipyridamole or BMP-2 *J. Biomed. Mater. Res. B Appl. Biomater* **105** 366-375

Jones A C, Arns C H, Sheppard A P, Hutmacher D W, Milthorpe B K and Knackstedt M A 2007 Assessment of bone ingrowth into porous biomaterials using MICRO-CT *Biomaterials* **28** 2491-504

Kerckhofs G, Schrooten J, Van Cleynenbreugel T, Lomov S V and Wevers M 2008 Validation of x-ray microfocus computed tomography as an imaging tool for porous structures *Rev. Sci. Instrum.* **79** 13711

Komlev V S, Peyrin F, Mastrogiacomo M, Cedola A, Papadimitropoulos A, Rustichelli F and Cancedda R 2006 Kinetics of In Vivo Bone Deposition by Bone Marrow Stromal Cells into Porous Calcium Phosphate Scaffolds: An X-Ray Computed Microtomography Study Tissue Eng. 12 3449-58

López A, Montazerolghaem M, Engqvist H, Karlsson Ott M, and Persson C 2014 Calcium Phosphate Cements with Strontium Halides as Radiopacifiers *J. Biomed. Mater. Res. B Appl. Biomater* **102** 250–59

Lovati A B, Lopa S, Recordati C, Talò G, Turrisi C, Bottagisio M, Losa M, Scanziani E and Moretti M 2016 In Vivo Bone Formation Within Engineered Hydroxyapatite Scaffolds in a Sheep Model *Calcif. Tissue Int.* **99** 209-23

Maazouz Y, Montufar E B, Guillem-Marti J, Fleps I, Öhman C, Persson C and Ginebra M P 2014 Robocasting of biomimetic hydroxyapatite scaffolds using self-setting inks *J. Mater. Chem. B* **2** 5378

Mastrogiacomo M, Komlev V S, Hausard M, Peyrin F, Turquier F, Casari S, Cedola A, Rustichelli F and Cancedda R 2004 Synchrotron Radiation Microtomography of Bone Engineered from Bone Marrow Stromal Cells *Tissue Eng.* **10** 1767-74

Montufar E B, Traykova T, Gil C, Harr I, Almirall A, Aguirre A, Engel E, Planell J A and Ginebra M P 2010 Foamed surfactant solution as a template for self-setting injectable hydroxyapatite scaffolds for bone regeneration *Acta Biomater.* **6** 876-85

National Research Council 1996 Guide for the Care and Use of Laboratory Animals *Wash. DC Natl. Acad. Press*

Otsu N 1979 A Threshold Selection Method from Gray-Level Histograms *IEEE Trans. Syst. Man Cybern.* **9** 62-6

Parkinson I H, Badiei A and Fazzalari N L 2008 Variation in segmentation of bone from micro-CT imaging: implications for quantitative morphometric analysis *Australas. Phys. Eng. Sci. Med.* **31** 160-4

Particelli F, Mecozzi L, Beraudi A, Montesi M and Baruffaldi F 2012 A comparison between micro-CT and histology for the evaluation of cortical bone: effect of polymethylmethacrylate embedding on structural parameters *J. Microsc.* **245** 302-10

Perilli E, Baruffaldi F, Visentin M, Bordini B, Traina F, Cappello A and Viceconti M 2007 MicroCT examination of human bone specimens: effects of polymethylmethacrylate embedding on structural parameters *J. Microsc.* **225** 192-200

Peyrin F 2011 Evaluation of bone scaffolds by micro-CT *Osteoporos. Int.* **22** 2043-8

Polak S J, Candido S, Levengood S K L and Wagoner Johnson A J 2012 Automated segmentation of micro-CT images of bone formation in calcium phosphate scaffolds *Comput. Med. Imaging Graph.* **36** 54-65

Ritman E L 2011 Current Status of Developments and Applications of Micro-CT *Annu. Rev. Biomed. Eng.* **13** 531-52

Roberts S J, Geris L, Kerckhofs G, Desmet E, Schrooten J and Luyten F P 2011 The combined bone forming capacity of human periosteal derived cells and calcium phosphates *Biomaterials* **32** 4393-405

Sheppard A P, Sok R M and Averdunk H 2004 Techniques for image enhancement and segmentation of tomographic images of porous materials *Phys. Stat. Mech. Its Appl.* **339** 145-51

Stalder A K, Ilgenstein B, Chicherova N, Deyhle H, Beckmann F, Müller B and Hieber S E 2014 Combined use of micro computed tomography and histology to evaluate the regenerative capacity of bone grafting materials *Int. J. Mater. Res.* **105** 679-91

Subramaniam S, Fang Y-H, Sivasubramanian S, Lin F-H and Lin C 2016 Hydroxyapatite-calcium sulfate-hyaluronic acid composite encapsulated with collagenase as bone substitute for alveolar bone regeneration *Biomaterials* **74** 99-108

Thimm B W, Wechsler O, Bohner M, Müller R and Hofmann S 2013 In Vitro Ceramic Scaffold Mineralization: Comparison Between Histological and Micro-Computed Tomographical Analysis *Ann. Biomed. Eng.* **41** 2666-75